Comparison of tagging in an educational context

Comparison of tagging in an educational context: Any chances of interplay?

Riina Vuorikari*

Open Universiteit Nederland (1),
P.O. Box 2960, 6401 DL Heerlen, NL

European Schoolnet, Rue de Treves 61,
1040 Brussels, Belgium
E-mail: Riina.Vuorikari@eun.org
*Corresponding author

Hans Põldoja

Tallinn University, Narva mnt 25,
10120 Tallinn, Estonia

University of Art and Design Helsinki,
Hämeentie 135c, FI-00560 Helsinki, Finland
E-mail: hans.poldoja@taik.fi
E-mail: hans.poldoja@tlu.ee

Rob Koper

Open Universiteit Nederland (1),
P.O. Box 2960, 6401 DL Heerlen, NL

Abstract: Web-based tagging systems for educational resources allow users to associate free keywords with resources. This paper first studies similarities and differences of tagging systems that are an integrated part of learning platforms, and then introduces a case study on teachers' use of social tagging tools. Analysing more than 20,000 teacher-created tags from five platforms, it was found that even if the tagging system designs differ, more than 30% of the posts were shared through tags between two or more platforms. Thus, tags can create interplay not only between the users and their tools, but also between diverse learning platforms.

Keywords: educational resources; tags; folksonomies; metadata; tagging systems; repositories; knowledge sharing; reuse.

Reference to this paper should be made as follows: Vuorikari, R., Põldoja, H. and Koper, R. (2010) 'Comparison of tagging in an educational context: Any chances of interplay?', *Int. J. Technology Enhanced Learning*, Vol.

Biographical notes: Riina Vuorikari is currently completing her PhD in OUNL. Her first degree is Master's in Education (University of Joensuu, Finland) and she received a diploma of Etudes Approfondis in Hypermedia (University of Paris 8, France). During her studies she has received several

stipends to participate in international exchange programmes. She has worked in European Schoolnet (EUN) since 2000 as a researcher. She deals with a variety of issues ranging from e-learning interoperability to the collaborative use of ICTs in schools. In the last years she has specialised in the use of social media in education.

Hans Põldoja is a research associate at the Institute of Informatics in Tallinn University and a doctoral student at the Media Lab in the University of Art and Design Helsinki. His research interests include collaborative authoring of open educational resources and design of learning environments. In recent years he has been responsible for the conceptual and interaction design in LeMill project.

Rob Koper is Professor and Dean of Centre for Learning Sciences and Technologies (CELSTEC) at the Open University of the Netherlands. He has more than 20 years of experience in the field and has numerous roles in management and advisory boards, like the National Assessment Agency, the Digital University Consortium, the local government, standardisation bodies like IMS, CEN/ISSS, elected member of the EU Prometeus board, and advisor for the new European Commissions' RTD programmes. He has published over 200 publications in scientific journal and books.

1 Introduction

End-user generated tags for learning resources in a Learning Object Repository (LOR) can be seen as part of the dialogue for the co-construction of knowledge and individual's participation in social interactions (Vygotsky, 1978; Engeström, 1987). Margaryan and Littlejohn (2008) suggest that Activity Theory offers a theoretical framework that allows studying LORs and communities as a single system, rather than as a loose set of instruments, subjects, objects and outcomes. Similarly, Activity Theory can be adapted to study LORs and tags as part of social interactions.

Using such a framework, the barriers and enablers of learning resource repositories' usage were studied (Margaryan and Littlejohn, 2008). One barrier was the mismatch between how repository curators and users perceived the services. The authors argue that curators' repository-centric perspective frequently leads to introduction of repositories as stand alone tools to users. However, they note that repositories are not used in isolation. They are part of the repertoire of tools that individuals and communities use to achieve learning goals. Therefore, they claim, the interplay between repositories and existing tools has to be considered.

Such an interplay with existing educational offers (e.g., learning resources, learning platforms, tools) is central to this research. In this contribution, the focus is on tagging tools that are offered as a feature of an educational LOR or platform, and how users (e.g., teachers) use these tools in an educational context. Seen from the user's point of view, a LOR is only one component within an entire repertoire of tools that she or he uses for a given information seeking task. Therefore, more flexibility for users is desirable. This paper focuses on the interplay between a number of educational resource platforms (Maier and Thalmann, 2008; Vuorikari and Põldoja, 2008).

First, a brief overview of the application domain, the learning resources repositories, is provided with an introduction to the fundamentals of social tagging. Section 3 introduces the educational resource platforms in this study, and using the taxonomy from Marlow et al. (2006), describes how tagging has been applied in these platforms and what are their differences and similarities. Then, a case study on users' learning resource collections is presented. Finally, a proof of concept for the interplay between five tagging systems in educational context is presented proposing that an aggregated cross-application tagcloud can potentially enhance the interplay between existing tools by offering novel ways of social navigation not only across applications, but also across language and national contexts. Section 6 discusses the results and future work, and lastly Section 7 offers a conclusion.

2 Learning resources, repositories and social bookmarking

Digital learning resources, and/or their associated metadata, are typically organised, classified and stored in online databases, often called LORs or digital libraries. A rich variety of LORs currently operate online facilitating learners', teachers' and tutors' access to learning resources in different contexts (e.g., disciplines, languages, curricula alignment) (Tzikopoulos et al., 2007). Previous studies have focused on the use of LORs in different educational contexts (McCormick et al., 2004; Strijker, 2004; Harley et al., 2006; Margaryan and Littlejohn, 2008; Petrides et al., 2008).

Conventionally, expert indexers or librarians catalogue learning resources using metadata standards such as the IEEE Learning Object Metadata (IEEE LOM, 2002) or the Dublin Core Metadata Initiative (2003). The quality of metadata can be evaluated from two different perspectives: its validity in terms of its ability to describe the resource and its usefulness for 'searchability' and how well it supports retrieval of the resource (Barritt et al., 2004).

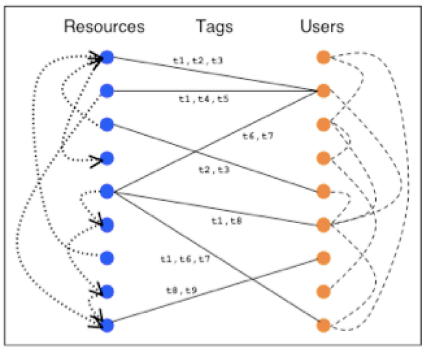
Searching learning resources both within a repository and across repositories using metadata is crucial (e.g., Ternier et al., 2008; Massart, 2009), as content often-times is not cross-referenced via hyperlinks. The situation is similar to 'Enterprise search' (Mukherjee and Mao, 2004), where content from heterogeneous repositories is made available on the intranet lacking the typical link-structure of the internet, and thus unsuitable for PageRank-type of algorithms (Brin and Page, 1998). LORs traditionally share a similar search problem: repositories from different institutions and countries offer content in similar curriculum areas, however, the content is seldom cross-referenced via hyperlinks. Especially in the European context, where learning resources are offered in a variety of national and regional settings (EdReNe, 2008), cross-referencing across national borders is rare. Halavais (2000) found that websites in most cases are more likely to link to another site hosted in the same country than to cross national borders. Language-wise, Berendt and Kralisch (2009) show that the smaller the language is, the smaller is the relative percentage of in-links. This indicates that non-English languages are under-represented on the web in terms of the links that content creators set in those languages, creating yet another barrier to cross-language searches.

In the last years, the proliferation of social media has not only changed how the production and use of metadata is perceived, but also the way users discover content through social networks. Social bookmarking services are a sub-group of social network sites. Boyd (2006) offers her definition: "social network site" is a category of websites

with profiles, semi-persistent public commentary on the profile, and a traversable publicly articulated social network displayed in relation to the profile. In bookmarking and tagging systems, each user has a profile, annotations are supported in terms of tags on the content artefacts, and the creation and support of implicit and explicit social networks emerge from different types of interactions. Tagging describes the act of end-users adding non-hierarchical, free keywords to any type of digital resource, regardless of its physical existence on a given service, repository, or database.

Marlow et al. (2006) propose a model for the underlying structure of the social tagging system with three main components; resources, tags and users. Figure 1 depicts this conceptual model where users assign tags to a specific resource; tags are represented as typed edges (i.e., links) connecting users and resources. Resources may also be connected to each other (e.g., as links between web pages) and users may be associated by a social network, or sets of affiliations (e.g., users that work for the same company). Moreover, Marlow et al. (2006) show that the socio-technical design of the system affects the information it generates. For the purpose of designing such systems they propose two taxonomies: "System design and attributes" which may affect the nature and distribution of tags and, therefore, the attributes of the information collected by the system; and 'User incentives' that largely affect how users behave, the forms of contribution allowed and the personal and social motivations for adding input to the system.

Figure 1 Model of a tagging system (see online version for colours)



Source: Marlow et al. (2006)

Tags and folksonomies thus offer complimentary ways to explore the content by searching, filtering, navigating, and exploring other users' tags and tagged items. Tagclouds, for example, consist of tags that users of a given tagging service have associated with the resources, and can be aggregated on different levels, such as personal, resource-specific or global tagclouds. According to Golder and Huberman (2006), tagging is like filtering; out of all the possible items that are tagged, a filter (i.e., a tag) returns only those items identified with that tag. From a user perspective, navigating a tag system is similar to conducting keyword-based searches; regardless of the implementation, users are providing salient, descriptive terms in order to retrieve a set of applicable items.

As the end-users add free keywords to photos, links and various other digital artefacts, the distinction between formal types of metadata and informal ways of 'tagging' resources starts to change. Critics suggest that social tagging and folksonomies are characterised by flaws that formal classification systems are designed to eliminate. In addition, social tagging and folksonomies all but invite deliberately idiosyncratic tagging, called 'meta noise', which burdens users and decreases the systems' information retrieval utility. It is argued that top-down taxonomies or ontologies enable more efficient indexing and searching of content (Guy and Tonkin, 2006).

Recent research on tags has focused on their use to enhance web search. Heymann et al. (2008) conclude that tags are unlikely to be much more useful than a full text search emphasising page titles, as over 50% of tags occur in the page's title, the body of text, or within backward or forward links. Yanbe et al. (2007), on the other hand, show that social bookmarks can be used to increase the precision and 'freshness' of a standard link-based search, and to extend the search capabilities of existing search engines. Santos-Neto et al. (2009) show that the item-based interest structure of social tagging is much more segmented than its tag-based counterpart, thus tag-based social browsing interface could be beneficial.

What makes social tagging systems different from conventional indexing approaches is the fact that they support and enhance social interactions. Such systems allow users to connect to other users, but also to their resources and tags. These connections happen through relationships that are formed between users, their resources and tags. These underlying social structures, or networks, become important for creating the missing cross-referencing structure that did not exist before between separate pieces of content. Millen et al. (2007) studied social bookmarking in IBM and suggest that when integrated with traditional search engines, social tagging has the potential to solve 'Enterprise search' problems.

Similarly, the goal is to study whether this could be the case for the domain of learning resources. Learning resource platforms such as LeMill, Learning Resource Exchange (LRE) and OER Commons have end-user tags and/or social bookmarking as part of the tool sets offered to users. In the following section the tags and the specifics of these different tagging systems in an educational context are studied to better understand their domain specific use. In addition, the similarities and differences between these systems and their outcomes, tags, are studied in order to investigate their further use for crossing different contexts (e.g., repository, language, nation/region) to enhance the discovery learning resources.

3 Comparison study of three tagging systems on learning resources

In order to investigate the possibilities of using tags as a way to access learning resources across different applications, three educational platforms are studied (Table 1). Early work on tagging systems in a general context has shown that systems have different dynamics; Marlow et al. (2006) find that system-level design choices and user incentives affect the nature and distribution of tags. Similarly, Sen et al. (2006) demonstrate that tag input systems affect the nature and distribution of tags, as well as the uptake of the tagging activity. Is this also the case with different educational tagging systems, and do tags differ from one platform to another?

Table 1 Dataset description of repositories

	Calibrate http://lreforschools.eun.org	*LeMill http://lemill.net*	*OER Commons http://www.oercommons.org*
1 Time span of data	November 2006–2007	May 2006–December 2007	December 2006-official release in March 2007
2 Number of resources tagged	682	3249	200
3 Number of tags	920	3543	244
4 Number of tags applied	1282	9257	502
5 Number of users tagging	142	436	91
6 Average of tags/resources	1.9 tags/resource	2.8 tags/resource	2.5 tags/resource
7 User incentives to tag*	"Keep found things found", personal retrieval	Share with groups; attract attention, future retrieval	Future retrieval, contribution and sharing
8 Objects types*	Textual, metadata of learning resources	User-contributed, self-authored resource (textual, non-textual)	User-contributed metadata of learning resources
9 Source of material*	System, from educational repositories	User	System, from educational repositories, users
10 Tagging rights*	Free-for-all tagging	Self-tagging, free-for-all	Free-for-all tagging
11 Tagging support*	Blind/viewable tagging	Blind tagging	Viewable tagging
12 Resource connectivity*	None	Grouped, linked	Linked
13 Social connectivity*	None	Grouped, linked	Grouped
14 Other annotations	Rating, comment (public/private)	Teaching/learning story	Ratings, my notes
15 Create a collection	Favourites = bookmark + tags	Collections, no tags related	myPortfolio, add tags possible
16 Language of tags	Multiple languages, users from five countries	Multiple languages, users from 39 countries	Mostly English, users from different countries

*Refers to the taxonomy of tagging systems proposed by Marlow et al. (2006).
The shaded cells indicate similarities between features.

3.1 Tagging systems and datasets

A uniform way to describe each resource platform with a tagging tool is used in this study to allow the comparison of their differences and similarities. First, a general description is given, then the user incentives to tag and the purposes for which the tags are used are explained for each platform. Following, the 20 most used tags in each system are sampled, their languages, frequency and number of users are studied. The tags are classified to factual, subjective and personal tags (Sen et al., 2006; Golder and Huberman, 2006). Then, a short discussion on tags in each system is provided.

The Calibrate portal, LeMill and OER Commons all share the same macro-context. They offer learning resources in the main curriculum area in English and in other

languages for the target audience of teachers, learners and educators. During the period when the datasets were acquired in December 2007, the Calibrate portal and LeMill mostly had users from the New Member countries in the EU, whereas OER Commons mostly served an American audience. The log-files were obtained from each repository for learning resources that contained at least one tag. The following data were provided for each record: user ID, resource ID and tag(s).

CALIBRATE portal (the current version is known as LRE http://lreforschools.eun.org)

The Calibrate portal provides federated search over a number of educational repositories that belong to European Schoolnet (2009) and its associated partners. The portal was only available to pilot schools, but has now been made available publicly. Users can search (free text and advanced) and browse educational resources through the portal and create their own collections of resources called 'Favourites'. Users can choose the interface language from ten different languages that can also be used for searching. During the time of the data gathering, the portal provided little collaboration through tags. In the current version, the tags are made public and sharable.

User incentives for tagging

Tagging on the Calibrate portal is related to the action of creating a bookmark for interesting learning resources that the user wants to 'keep found'. Users can thus create their own collections of resources to access them later.

Purpose of tags

The purpose of tags on the Calibrate portal is personal and facilitates individual's future retrieval of interesting resources previously found on the portal. In other words, a user is related to his own collection of resources through personal tags. Tags were also used for free-text search.

Discussion on tags

Table 2 lists the most used tags in the Calibrate portal. Sharing through tags mostly happened though the top three tags that were applied relatively often by a group of about ten users. Additionally, some coincidental sharing takes place: the tag 'Matematika' is shared by both Hungarian and Czech users because the same word is used in both languages. Apart from these cases, there is little sharing of tags among users. The low number of users who applied tags can be observed, e.g., one or two users have created collections of resources (e.g., chemistry or geometry).

Low sharing of tags among users is most likely due to the design decision of semi-blind tagging and the fact that tags are not displayed to other users. The low sharing of tags amounts to little convergent among tags and little emerging folksonomy can be observed among users. A manual analysis of the global tags in the Calibrate system reveals that 90% of them have been applied only once by one user. Looking at the tags, some tags are found that, even if not translated, can be rather easily understood thanks to their similar spelling in many languages (e.g., literature, fyzika, chemie, europa, evropa). Tags that hardly even need translation (e.g., test) were identified. These tags are loosely grouped under the umbrella of 'travel well' tags, as they propose added value for multilingual users (Vuorikari and Ochoa, 2009). From the same study, it was found that tags used in Calibrate are mostly of a factual type; they identify properties of the objects such as the topical area of the resource and some other attributes, seldom any qualitative properties. This trend is also visible in Table 2.

Table 2 Twenty most used tags in calibrate. Translations provided in (). ISO 639-1 language codes are used for languages, as in all the tables hereafter

Calibrate	*Language*	*Number of applications*	*Tag category*	*Users*
Külföldi jó (good foreign resource)	hu	70	Subjective	10
Külföldi közepes (average foreign resource)	hu	52	Subjective	10
külföldi gyenge (weak foreign resource)	hu	37	Subjective	11
Angličtina (English)	cz	20	Factual, topic	3
Értékeltek (evaluation)	hu	16	Subjective	2
Literatura	cz	14	Factual, topic	2
Matematika	cz, hu	13	Factual, topic	7
Global problems	en	11	Factual	2
Test	'travel well'	10	Factual, type	3
Vesmír (Space)	cz	7	Factual, topic	2
Fyzika	cz	7	Factual, topic	2
Chemie	cz	6	Factual, topic	1
English in general	en	6	Factual, topic	2
Europa	'travel well'	6	Factual, country	2
Geometrie	hu	6	Factual, topic	1
Fénytan (optics)	hu	5	Factual, topic	1
Animáció	hu	5	Factual, type	4
Evropa	'travel well'	5	Factual, country	3
Planety (planets)	pl	5	Factual, topic	2
Safety	en	5	Factual, topic	2

LeMill (http://lemill.net)

LeMill is a web community for finding, authoring and sharing learning resources. It is divided into four sections: Content, Methods, Tools and Community. The main target audience are primary and secondary school teachers, but anyone can join. Registered users can publish learning content and descriptions of educational methods and tools. It is a wiki-like system where all the learning resources are published under an open licence and can be edited by other members.

User incentives for tagging

Tagging in LeMill is part of the content authoring workflow that includes creating the resource, adding metadata and publishing the resource. Tags are not a mandatory metadata element. The main motivation for the content creator to add tags is sharing the resource with other users. Second, tags help attract attention to a creator's content through the tagcloud, which has a central role in the navigation. Last, content creators can use tags as a personal management tool to keep their own resources organised. Personal tagclouds can be accessed through the user's profile. Separate from tagging is a tool that allows users to create personal collections of resources. Content (learning resources, media pieces, references), methods and tools can be added to a collection to easily access them later and share them with others. It is not possible to add tags for collections.

Purpose of tags

The main purpose of tags is to be visible in a tagcloud, one of the main navigation tools. Similar cloud-like navigations have been created around other metadata too, like language, subject area and intended audience. Tags are also a way to contribute to the system and share resources among groups.

Discussion on tags

An example of sharing through tags is shown in Table 3 (e.g., like calibrate, r, lemill and dlf07tallinn, the latter tag stands for "deer leap foundation, tallinn, 07"). These are tags decided by a community that allows sharing the resources and to aggregate a thematic collection around a tag. Even if these tags are powerful for sharing and retrieving resources among a given group, they are less descriptive for the global audience.

Table 3 also reveals less-formal groups or ad-hoc communities that have been formed around some resources (e.g., matemaatika and matematika). These tags can also be 'travel well' tags, as they are shared by different language communities. Seen from the small number of users for some tags, it can be inferred that tags in LeMill also are used for personal management to create personal collections, e.g., 'projektijuhtimine' (2 users), 'hambad' (2 users), geomeetria (1 user) and 'felvilagosodas' (1 user).

Table 3 Twenty most used tags in LeMill. Translations provided in ()

LeMill	*Language*	*Number of applications*	*Tag category*	*Users*
Calibrate	–	136	Personal, shared	36
r	–	116	Personal, shared	4
Algebra lineal (Linear algebra)	sp	97	Factual, topic	19
Projektijuhtimine (Project management)	et	82	Factual, topic	2
Matemaatika	et	69	Factual, topic	16
Lemill	–	65	Personal, shared	15
Kõneravi (Pronaunciation problems)	et	64	Factual, topic	6
A first course in linear algebra	en	54	Factual, topic	10
Hambad (Teeth)	et	49	Factual, topic	2
Algebra	'travel well'	48	Factual, topic	7
Matematika	cz, hu, lt	47	Factual, topic	12
Geomeetria	et	46	Factual, topic	1
Traduccion	sp	44	Factual, topic	8
Felvilágosodás (enlightenment)	hu	38	Factual, topic	1
Linnud (birds)	et	38	Factual, topic	16
Логика (Logics)	ru	38	Factual, topic	21
Algõpetus	et	38	Factual, topic	2
dlf07tallinn	–	37	Personal, shared	1
English	en	37	Factual, topic	21

OER Commons (http://www.oercommons.org)

The OER Commons allows users (teachers and professors from pre-K to graduate school) to access and share course materials and learning resources that are harvested from a number of collaborating educational repositories around the world, as well as added by users. Anyone can access resources, a number of search features that have been made available (text, advanced search, browsing topics and tags). Additionally, authenticated users are offered more features such as creating their own collections, adding tags and sharing their material with other users.

User incentives for tagging

The OER Commons encourages users to add searchable metadata, such as tags, to create user's personal keyword vocabulary. The motivation for tags is similar to what Marlow et al. (2006) call "Contribution and sharing: to add to conceptual clusters for the value of either known or unknown audiences". Additionally, users can create private collections in 'MyPortfolio'. In this area the user can see all the saved items as a list with associated tags on the site. The user cannot, however, use the tags to filter these resources.

Purpose of tags

The OER Commons focuses on providing tags as additional metadata that users can use to access resources. Tags, when displayed next to conventional metadata of the resource description, can give additional cues to other users on the content and its use by creating a third-party conceptual cluster of tags. Tags also support discovery of resources; there are both a system and resource level tagclouds for navigation.

Discussion on tags

From Table 4 it can be seen that some tags are used by many users (e.g., algebra, evolution and education) indicating a small community forming around the topic. There are also tags that are used clearly only for personal collections of resources (e.g., flu, urban). These provide both added value for the other users through the tagcloud and resource-specific tags. Tags in Table 4 are all factual, the type of tag which adds high value to other users (Sen et al., 2006). Additionally, tags are all in English, which indicates that most users either have English as their mother tongue or use English to facilitate sharing.

Table 4 Twenty most used tags in OER Commons

OER Commons	*Language*	*Number of applications*	*Tag category*	*Users*
Biology	en	20	Factual, topic	15
Art	en	11	Factual, topic	3
Globalisation	en	10	Factual, topic	3
Psychology	en	10	Factual, topic	3
Media	en	9	Factual, topic	4
Internet	'travel well'	9	Factual, topic	5
Writing	en	8	Factual, topic	4
Science	en	8	Factual, topic	6
Civil society	en	8	Factual, topic	3

Table 4 Twenty most used tags in OER Commons (continued)

OER Commons	*Language*	*Number of applications*	*Tag category*	*Users*
Flu	en	7	Factual, topic	1
Education	en	7	Factual, topic	6
Evolution	en	7	Factual, topic	6
Urban	en	7	Factual, topic	1
Engineering	en	7	Factual, topic	4
Algebra	en	6	Factual, topic	6
E-learning	en	6	Factual, topic	2
Environment	en	5	Factual, topic	3
Chemistry	en	5	Factual, topic	4
Research	en	5	Factual, topic	3
French	en	5	Factual, topic	3

3.2 Differences and similarities between tagging systems and tags

3.2.1 Differences

There are a number of differences between the tagging systems. They can be observed in Table 1 which uses the taxonomy of characteristics of system design and user incentives. Similarities in each category are highlighted using grey cells. The main difference is in the logic of the tagging system, which is also related to the incentive for tagging: In Calibrate, tags are purely for personal retrieval purposes (Favourites), whereas in LeMill tags have the purpose of attracting other users (tagcloud) and sharing resources. In OER Commons, on the other hand, tags are searchable, additional metadata.

The tagging rights and types of objects to tag also vary; LeMill is a clear example of self-tagging (e.g., Flickr), where the type of object being tagged is typically a resource or a reference created by the user. In Calibrate and OER Commons, users mostly tag resources that are created by someone else. Users actually tag only the metadata reference of the resources, which might reside on some other educational repository.

Users tag differently (Table 5); 80% of users in Calibrate have only applied one tag to a resource, whereas in LeMill and OER Commons, users apply more tags. In LeMill, where the creator of the resource mostly adds tags, about 75% of resources have two or more tags and in OER Commons about 60% of resources have more than one tag.

Table 5 Percentage of tags per resource in each platform

Tags/resource	*LeMill* (%)	*OER Commons* (%)	*Calibrate*
1 tag	28	41	80%
2 tags	22	18	15%
3 tags	20	22	3 more 5%
4 or more tags	30	19	–

Looking at the nature of tags in each system, it can be seen that in OER Commons tags are very factual. This can be due to visibility of tags (tagclouds and related

tags are displayed). Due to common workshops and project related activities more subjective tags can be observed in Calibrate (Table 2, No: 1, 2, 3) and LeMill (Table 4, No: 1, 2, 5). Both OER Commons and LeMill have more convergent folksonomies starting to emerge, whereas in Calibrate, sharing the same vocabulary between users of the system is less. Lastly, differences in the languages in which people tag are observed: the most used tags in OER Commons are in English, whereas in Calibrate and LeMill different languages reflect the user base of each system.

In addition to tagging, LeMill and OER Commons offer a different tool, namely *Collections*, that allow users to easily access and share interesting resources. When comparing this knowledge to the model of social bookmarking system presented above, we understand that in these three systems, not only are tags important for creating the underlying connections between users, tags and resources, but it is useful to consider resources that in *Collections*, as they clearly indicate that users have shown interest in these resources.

3.2.2 Similarities

Although there are many differences in design decisions on the system level, the purpose of tags in each system, and the incentive schemes for users to tag, in the sample of the most used tags (20). It can be noted that they are very similar in their nature. A majority of them are factual, and represent properties that might be useful for other users of different educational systems.

A manual comparison of the most used tags was conducted on a pairwise basis, as in Muller (2007). It was found that common tags appear in each pair of the tagging services; Calibrate-LeMill (8), LeMill-OER (8) and Calibrate-OER (4). These tags contain semantic similarities: they share the same tag (e.g., algebra), cover similar topical areas (e.g., biology, Birds, linnud=birds in Estonian) or the same topic in different languages (e.g., chemistry, chemie). 18 out of 60 tags appeared in more than one service.

Inspired by these similarities, the overlap of tags in all three services was studied by analysing 4707 distinct tags in the datasets. Table 6 shows that 19 tags were shared among all three services (0.4% of all distinct tags) resulting in 313 tag applications (2.7% of all posts). These tags are listed in Table 7, where it can be seen that similarities were not only found among the most popular tags, but also in the 'long tail', i.e., among tags that had been applied only a few times. It was also found that about 5% of the distinct tags are shared between two services which results to 14.2% of all tag applications on these platforms, thus forming link-structures across learning resource platforms through tags (pair 'tag-item').

Table 6 Tags that appear in more than one service

n = 4707 distinct tags	*Number of distinct tags*	*Percentage of all distinct tags* (%)	*Tag applications* (*percentage of all*)
Tags shared in 3	19	0.4	313 (2.7%)
Tags shared in 2	244	5.2	1654 (14.2%)

A notable similarity between tags in each system is that they cover a number of the topical areas that are shared among many of the educational systems (e.g., mathematics, sciences). Moreover, 'travel well' tags were found in each repository. These tags can be

found useful thanks to their similarity in spelling in many languages. These are names such as 'internet', place names (e.g., Europe), and commonly known acronyms (e.g., eu). They are easily understandable in many languages and do not always need to be translated, thus they are powerful in a multilingual context.

Table 7 Tags that appear in all three platforms for learning resources

Tag	*Number of applications*
Algebra	58
Biology	37
Internet	27
Europe	23
Environment	22
Water	21
Art	19
Music	15
Grammar	14
Education	13
Film	13
Science	12
London	9
Eu	7
German	6
Culture	5
Climate	4
Games	4
Quiz	4

3.3 Conclusion and discussion

In this section the differences and similarities of tagging systems for educational resources were studied. It can be concluded that the tagging systems in an educational context can be described using the common taxonomy for social bookmarking systems. When positioned on the dimensions of the tagging design taxonomy by Marlow et al. (2006), the educational tagging tools represent rather different system types, almost similarly to the comparison that the same authors made on delicious.com vs. Flickr. Tags produced by end-users in these different tagging systems appear very similar despite big differences in system-level design choices and user incentives. The similarities most likely stem from the similarities in the context and the user-base (e.g., teachers), who mostly teach similar curriculum areas (i.e., macro-contexts) despite differences in national and regional curriculum and standards alignment, rather than the inherent differences in the tagging systems as explained by Marlow et al. (2006).

In the next Section we look at how teachers use multiple information seeking strategies, as they are rarely satisfied with the offering of a single educational platform or provider.

4 Social bookmarks on non-institutionalised collections of learning resources

Teachers use a plethora of ways to discover educational content online. Harley et al. (2006) report on search strategies of 4500 US faculty members where Google-like searches are, by far, the most prominent (81%), second most important being personal Collections of resources (72%), and followed by 'portals' that provide links to disciplinary topics (55%). In a user group comprised of 45 language and science teachers in K-12 education, such diversity of strategies was also observed: one third use national and regional educational repositories as their primary source of educational content, 28% use search engines, 21% said they create their own content, 7% use content from schoolbook publishers and 12% reported all of the above.

These different search strategies highlight the argument from Margaryan and Littlejohn (2008) that learning resource repositories are not used in isolation; rather, a diversity of tools is deployed. In the following, a case study is presented on 16 teachers who have an account on both the LRE portal (continuation of Calibrate, hereafter the portal) and on *delicious.com* to study how different types of social bookmarking tools are used.

4.1 Method

The users are primary and secondary teachers in science, language learning and ICTs from Finland, Estonia, Hungary and Belgium. Seven are female and nine male. One participant is under 30 years old, eight are between 30 and 40 years, five between 40 and 50 years, and two are between 50 and 60 years old. Most of the participants were first introduced to *delicious.com* during the MELT summer school in 2007. In March 2008 they were invited to create a profile on the portal, which collects attention metadata regarding the learning resources bookmarked on the portal (posts). This includes information about the resource itself (e.g., LOM) and the tags applied. We asked for their *delicious.com* usernames to be part of this study when they participated in the MELT Summer school in 2008, where additional user observations and interviews were conducted (Zens, 2009).

From *delicious.com*, using the html-export service, users' 100 last posts, including the tags were gathered. The total number of posts was recorded, as well as all the tags applied and usernames within the network. Table 8 presents the data sets; the term 'distinct' for a tag or a resource that has been recorded in the system is used, as opposed to applied, which means how many times the tag has been associated with a post or how many times the same resource appears in collections.

Table 8 The data sets from *delicious.com* and the MELT portal

	Users	*Posts downloaded*	*Distinct resources*	*Distinct tags*	*Tag applications*	*All posts from this Group*
delicious.com	16	1176	1081	944	1583	3222
LRE portal	16	245	107	301	665	245

4.2 Results

A manual analysis of the 50 most used distinct tags associated with posts in *delicious.com* was performed to assess the nature of these resources. Almost all of them were related to educational context, such as teaching in general and, often, teaching English and grammar. Additionally, the URLs were analysed to check whether they contain certain keywords (e.g., esl, English) and names (e.g., wiki, blog, YouTube, LeMill, Sulinet). Table 9 shows a sample of posts that matched with the keywords, they comprise 57% of the downloaded posts. 52% of these posts are somehow related to learning resources, such as learning resource portals, science and language learning websites, dictionaries and reference material. 40% of the URLs indicate user-generated content such as blogs, wiki pages, Google-pages, photos and YouTube, also often related to educational activities. 7% seem to point towards software for the web and media.

Table 9 Type of content that is found in teachers' *delicious.com* accounts

Type of content in teachers' delicious.com accounts	*Number of posts*	*Percentage*
Learning resources (e.g., portals, science and language learning resources, maps)	353	52
User generated content (e.g., wikis, blogs, photos, Slideshare, YouTube)	273	40
Other software tools for media and web	49	7
Percentage of all downloaded posts	*675*	*57*

Moreover, Table 7 shows that the amount of posts in delicious.com by the studied group is substantial (3222): the median amount of posts was 105.5 per account compared to 15 per account in LRE. 59% of delicious.com users were above average, which can indicate a dedicated and systematic use of the tool. Interestingly, there seemed to be very little overlap of resources in users' collections in delicious.com. Whereas in LRE, the number of distinct resources is more than twice as high as the posts (245 vs. 107), in *delicious.com* this number is almost equal (1176 vs. 1081).

4.3 Discussion

Although the data sets are not directly comparable (most users have been using *delicious.com* more or less for a year, whereas the portal only for three months), they point in the same direction as the previous research (Harley et al., 2006; Margaryan and Littlejohn, 2008): teachers apply multiple strategies and use different sources to gather online teaching material, both from institutional sources like LORs by national educational authorities, but also other sources, as well as user generated content. It also raises the question whether repositories and educational platforms should integrate social tagging tools to their services, which in the best case allows communities to form around their content, but at worst, sets the limits only to the offering from the given platform. Or should the use of existing tools like *delicious.com* or *Diigo* be encouraged? In the following section, some reflection on this question is offered by introducing a mid-way solution.

5 Sharing tags across educational tagging systems

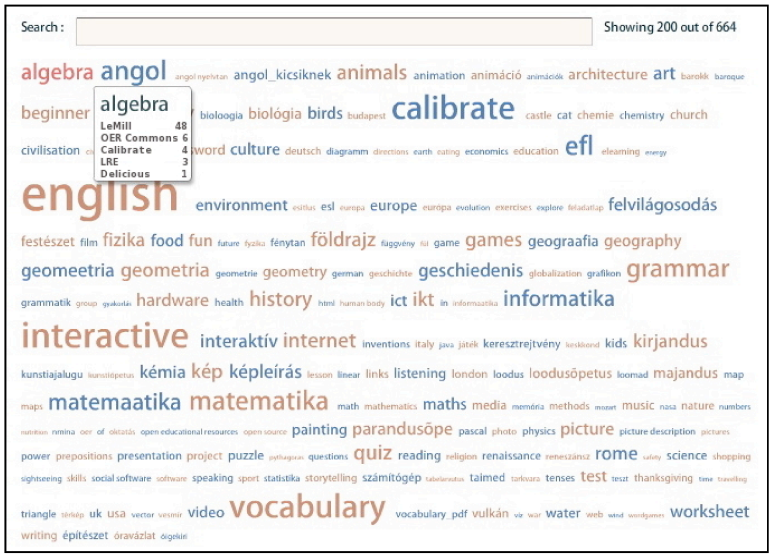
It has been demonstrated that the tags in different educational platforms share strong similarities. Instead of sharing only resources and their respective Learning Object Metadata, sharing metadata such as 'tag-item' pair becomes interesting. Secondly, in the case study, teachers' use of multiple platforms to search for suitable learning resources was studied. Inspired by an observation that curators' repository-centric perspective frequently leads to introduction of repositories as stand-alone tools to users, interplay between a number of tools (e.g., repositories, authoring and collaboration platforms, social bookmarking services) that individuals and communities use in educational context seems a desirable solution.

All the tags from different services in the previous dataset were collected and the same process of analysing overlap between tags was followed. This results in 9036 distinct tags from Calibrate, LeMill, OER Commons, LRE Portal and delicious.com. It was found that 666 of the distinct tags (7.4% of all distinct tags) overlap at least in two out of five different tagging systems (Table 10). They result in 6452 tag applications, which cover 30% of all the posts in the dataset. Using this 'tag-item' relation, an aggregated tagcloud can be created which comprises tags that are shared with two or more tagging systems in educational context (Figure 2).

Table 10 Tags shared among five different tagging tools in an educational context

Tags appear	*Distinct tags*	*Applications*	*Percentage of tag applications*
In more than 2 platforms	147	3047	16.0
In 2 platforms	519	3405	14.3
Total in 2 or more platforms	666	6452	30.3
All tags 5 services	*9036*	*21269*	

Figure 2 An aggregated tagcloud for learning resources from five different tagging applications (see online version for colours)



In Figure 2 the tag 'algebra' is highlighted. The user sees that this tag has been used by other users in different learning resource platforms: LeMill (48), OER Commons (6), Calibrate (4), LRE (3) and once in delicious.com. By clicking on any of the names, the user is taken to the respective platform and its search interface, where a list of these handpicked and tagged resources is found. Almost seamlessly to the user, she has crossed the system border to another resource platform and finds resources that users in that given community have indicated suitable to be used for 'algebra'. This type of aggregated 'cross-platform tagcloud' creates novel, community-based social navigation systems that take advantage of users' participation in social interaction and co-construction of knowledge. Such a tagcloud could be offered by each platform in addition to their other search tools.

6 Discussion and future work

In this paper the similarities and differences among three different tagging systems for educational resources has been studied. Additionally, a case study on educators' use of tagging both on a resource portal and using a generic tagging tool to manage their private collections of learning resources was presented. It was shown that even if the tagging system design decisions differ, the outcomes, i.e., the tags, are very similar across applications. This is allocated to similar macro-contexts: users (e.g., educators, learners) have interest in similar discipline areas and share a number of similar learning and teaching tasks, even across languages, curriculum and national contexts. Moreover, it has been shown that the interplay between the tools and users could be created through an aggregated cross-application tagcloud. Yet another form of interplay emerges, namely that of content coming from heterogeneous repositories that typically do not cross-reference each other via link-structures.

The idea of allowing users to access resources originating from different platforms through tags is complimentary to other forms of sharing learning resources and their metadata between repositories (e.g., Ternier et al., 2008; Massart, 2009). The proposal of a cross-platform tagcloud, though, introduces three new aspects. First, it builds on the social interactions among users in terms of co-construction of knowledge as tags, and secondly, it uses them as a way to offer interplay between learning resource platforms. Lastly, it introduces the idea of accessing both institutional resources (usually subject to some quality control within a closed information retrieval system) and private collections of resources from various sources. Such ideas are novel in the area of LOR, where the *de facto* way of sharing resources is based on federating and harvesting metadata. Instead of accessing the entire set of 'conventional' metadata, which can amount to thousands of resources (e.g., the LRE alone makes more than 35,000 resources available), 'human-made' filters, i.e., tags, bridge between platforms and guide the user's choice of resources.

Positive feedback has been received regarding the use of tagclouds by teachers in a federation of repositories (Zens, 2009). However, "fit for purpose" regarding users' information seeking tasks is important. Sinclair and Cardew-Hall (2008) show that where the user's information-seeking task was more general, participants preferred the tagcloud, but, when the information-seeking task required specific information, participants preferred the search interface. Ways to liase between search tasks, contexts and different ways to search is important. Additionally, future work in this area should concentrate

on assessing the intellectual value of tags (e.g., Farooq et al., 2007), as well as multilinguality of tags (Vuorikari and Ochoa, 2009).

Making the user experience more coherent and flexible through the integration across applications, rather than creating one monolithic system that is expected to be used by all, can play an important role. As in the social software scene, where users are offered tools to track their participation on diverse applications (e.g., APML, ULML), similar tools could be offered for learners and teachers to keep track of their attention and participation (e.g., content and communication in a large sense) across educational applications (Vuorikari, 2008). This, however, requires efforts from the educational application providers, for example, to generate metadata regarding users' attention in web feeds (e.g., Rss, Atom). Interoperability and data portability, not to mention the privacy, become crucial for the reuse of data.

Lastly, it has been shown that content coming from heterogeneous repositories that typically do not cross-reference each other via link-structures has such cross-references thanks to the triple (user, item, tag). Therefore, the link-structures from the aggregated tagcloud open more sophisticated avenues for resource discovery across contexts (e.g., language, country, curriculum, repository). Future work focusing on using these underlying connections to create measures of resources' importance will offer plenty of research challenges. Similarly to the Page-Rank algorithm (Brin and Page, 1998), tags, creating underlying connections between seemingly random pieces of content in different languages (and from repositories in different countries), rely on humans' subjective idea of their importance for a given information-seeking task. Using this new, emerging link-structure, and involving tags as 'anchor texts' (Kinsella et al., 2008), could offer totally new ways to "organise the world's learning resources and make them universally accessible and useful", similar to what Google claims its mission statement is for world's information. Additionally, the resource's potential for crossing across different contexts could be detected from the same link-structure. Resources-specific tags, for example, that appear in many different languages could indicate that the resource is being used in different language contexts and thus has potential to be used across contexts. Similarly, resources with users from a number of different countries could indicate that these resources are being used in different country and curriculum contexts. Conversely, resources that have tags associated to them only in one language or only by users from the same country as the resource is, could be disregarded and given less importance for the across-context discovery.

7 Conclusions

In this paper, tagging systems and the interplay between users and tools was studied, and on the other hand, the focus was on tags and resources. It has been demonstrated how the end-user generated tags create cross-references between separate pieces of content which reside in heterogeneous content platforms in a multilingual context. The triple (user, item, tag) helps create novel link-structures between cross-language content and offers new ways to take advantage of the methods known in the field of social information retrieval (e.g., social navigation, ranking of resources and social recommendations) in a multilingual context. The analyses in this paper lay the groundwork for social search ecology between a more conventional and formal metadata schemas, and user generated tag-based interest structures to allow novel ways to discover

learning resources (both content and other users) across repositories, languages and across national and regional curriculum.

Acknowledgements

The authors would like to thank their colleagues at OER Commons, LeMill and European Schoolnet for the tag exports, and wish for further collaboration in the future. They would also like to express their thanks to ManyEyes for their visualisation service, and to Matti for proofreading the paper.

References

Barritt, C., Alderman, F. and Lee, J. (2004) *Creating a Reusable Learning Objects Strategy: Leveraging Information and Learning in a Knowledge Economy*, Pfeiffer, California, USA.

Berendt, B. and Kralisch, A. (2009) 'A user-centric approach to identifying best deployment strategies for language tools: the impact of content and access language on web user behaviour and attitudes', *Inf. Retr*., Vol. 12, No. 3, pp.380–399.

Boyd, D. (2006) *Social Network Sites: My Definition*, Retrieved 23 January, 2007, from http://www.zephoria.org/thoughts/archives/2006/11/10/social_network_1.html

Brin, S. and Page, L. (1998) 'The anatomy of a large-scale hypertextual web search engine', *Computer Networks and ISDN Systems*, Vol. 30, pp.107–117.

Dublin Core Metadata Initiative (2003) ISO 15836: http://www.dublincore.org

EdReNe (2008) *EdReNe: Current State of Educational Repositories – National Overview*, Project Deliverable, Aarhus, Denmark, Uni-C, Retrieved 14 July, 2009, from http://edrene.org/results/currentState/index.html

Engeström, Y. (1987) *Learning by Expanding: An Activity Theoretical Approach to Developmental Research*, Orienta-Konsultit Oy, Helsinki, Retrieved from http://lchc.ucsd.edu/MCA/Paper/Engestrom/expanding/toc.htm

European Schoolnet (2009) Retrieved 17 July: from http://www.eun.org

Farooq, U., Kannampallil, T.G., Song, Y., Ganoe, C.H., Carroll, J.M. and Giles, L. (2007) 'Evaluating tagging behavior in social bookmarking systems: metrics and design heuristics', *Proceedings of the 2007 International ACM Conference on Supporting Group Work*, 4–7 November, Sanibel Island, USA, pp.351–360.

Golder, S. and Huberman, B.A. (2006) 'The structure of collaborative tagging systems', *Journal of Information Science*, Vol. 32, No. 2, pp.198–208.

Guy, M. and Tonkin, E. (2006) *Tidying Up Tags*, *D-Lib Magazine*, Vol. 12, Month 1, http://www.dlib.org/dlib/january06/guy/08guy.html

Halavais, A. (2000) 'National borders on the World Wide Web', *New Media and Society*, Vol. 2, No. 1, pp.7–28.

Harley, D., Henke, J., Lawrence, S., Miller, I., Perciali, I. and Nasatir, D. (2006) *Use and Users of Digital Resources: A Focus on Undergraduate Education in the Humanities and Social Sciences*, Available from: http://cshe.berkeley.edu/research/digitalresourcestudy/report/digitalresourcestudy_final_report.pdf

Heymann, P., Koutrika, G. and Garcia-Molina, H. (2008) 'Can social bookmarking improve web search?', *Proceedings of the International Conference on Web Search and Web Data Mining*, ACM, Palo Alto, California, USA, pp.195–206.

IEEE LOM (2002) 'Standard for learning object metadata', *IEEE Learning Technology Standards Committee*, IEEE 1484.12.1.

Kinsella, S., Budura, A., Skobeltsyn, G., Michel, S., Breslin, J.G. and Aberer, K. (2008) 'From Web 1.0 to Web 2.0 and back: How did your grandma use to tag?', *Proceeding of the 10th ACM Workshop on Web Information and Data Management*, ACM, Napa Valley, California, USA, pp.79–86.

Maier, R. and Thalmann, S. (2008) 'Institutionalised collaborative tagging as an instrument for managing the maturing learning and knowledge resources', *International Journal of Technology Enhanced Learning*, Vol. 1, Nos. 1/2, pp.70–84.

Margaryan, A. and Littlejohn, A. (2008) 'Repositories and communities at cross-purposes: issues in sharing and reuse of digital learning resources', *Journal of Computer Assisted Learning (JCAL)*, Vol. 24, No. 4, pp.333–347.

Marlow, C., Naaman, M., Boyd, D. and Davis, M. (2006) 'HT06, tagging paper, taxonomy, flickr, academic article, to read', *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia*, Odense, Denmark, pp.31–40.

Massart, D. (2009) 'Towards a pan-European learning resource exchange infrastructure', in Feldman, Y., Kraft, D. and Kuflik, T. (Eds.): *Proceedings of the 7th conference on Next Generation Information Technologies and Systems (NGITS'2009)*, LNCS 5831, Haifa, Israel, Springer, pp.121–132.

McCormick, R., Scrimshaw, P., Li, N. and Carmel, C. (2004) *Celebrate Evaluation, Deliverable 7.2. Brussels*, European Schoolnet, Available from http://celebrate.eun.org/eun.org2/eun/Include_to_content/celebrate/file/Deliverable7_2EvaluationReport02Dec04.pdf

Millen, D., Yang, M., Whittaker, S. and Feinberg, J. (2007) 'Social bookmarking and exploratory search', in Bannon, L., Wagner, I., Gutwin, C., Harper, R. and Schmidt, K. (Eds.): *ECSCW'07: Proceedings of the Tenth European Conference on Computer Supported Cooperative,* Limerick, Ireland, pp.24–28.

Mukherjee, R. and Mao, J. (2004) 'Enterprise search: tough stuff', *Enterprise Search*, Vols. 2, 4, pp.36–46.

Muller, M.J. (2007) 'Comparing tagging vocabularies among four enterprise tag-based services', *Proceedings of the 2007 International ACM Conference on Supporting Group Work*, 4–7 November, Sanibel Island, USA, pp.341–350.

Petrides, L., Nguyen, L., Jimes, C. and Karaglani, A. (2008) 'Open educational resources: inquiring into author use and reuse', *International Journal of Technology Enhanced Learning*, Vol. 1, Nos. 1/2, pp.98–117.

Santos-Neto, E., Condon, D., Nazareno, A. and Ripeanu, M. (2009) 'Individual and social behavior in tagging systems', *Proceedings of the 20th ACM Conference on Hypertext and Hypermedia*, Torino, Italy, pp.183–192.

Sen, S., Shyong K.L., Cosley, D., Rashid, A.M., Frankowski, D., Harper, F. *et al.* (2006) 'Tagging, community, vocabulary, evolution', *Proceedings of CSCW 2006*, Banff, Canada, pp.181–190.

Sinclair, J. and Cardew-Hall, M. (2008) 'The folksonomy tagcloud: When is it useful?', *J. Inf. Sci*., Vol. 34, No. 1, pp.15–29.

Strijker, A. (2004) *Reuse of Learning Objects in Context: Technical and Human Aspects*, PhD Thesis, University of Twente, Netherlands.

Ternier, S., Massart, D., Campi, A., Guinea, S., Ceri, S. and Duval, E. (2008) 'Interoperability for searching learning object repositories: the ProLearn query language', *D-Lib Magazine*, Vol. 14, Nos. 1/2.

Tzikopoulos, A., Manouselis, A. and Vuorikari, R. (2007) 'An overview of learning object repositories', in Northrup, P. (Ed.): *Learning Objects for Instruction: Design and Evaluation*, Idea Group Publishing, Hershey, PA, pp.29–55.

Vuorikari, R. (2008) 'A case study on teachers' use of social tagging tools to create collections of resources – and how to consolidate them', in Wild, F., Kalz, M. and Palmer, M. (Eds.): *Proceedings of the First International Workshop on Mashup Personal Learning Environments*, CEUR, Maastricht, Netherlands, pp.65–68.

Vuorikari, R. and Ochoa, X. (2009) 'Exploratory analysis of the main characteristics of tags and tagging of educational resources in a multi-lingual context', *Special Issue of Social Information Retrieval for Technology Enhanced Learning*, *Journal of Digital Information*, Vol. 10, No. 2.

Vuorikari, R. and Poldoja, H. (2008) 'Comparison of educational tagging systems – any chances of interplay?', *Proceedings of the 2nd SIRTEL'08 Workshop on Social Information Retrieval for Technology Enhanced Learning*, CEUR, Maastricht, Netherlands, p.382.

Vygotsky, L.S. (1978) *Mind in Society: The Development of Higher Psychological Processes*, Harvard University Press (the original in Russian published in 1930), Cambridge, MA.

Yanbe, Y., Jatowt, A., Nakamura, S. and Tanaka, K. (2007) 'Can social bookmarking enhance search in the web?', *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*, ACM, Vancouver, BC, Canada, pp.107–116.

Zens, B. (2009) *MELT Deliverable 7.3 Final Evaluation Report*, Brussels, European Schoolnet, Available from: http://info.melt-project.eu/shared/data/melt/MELT_D7_3_Final_Evaluation_Report.pdf

Websites

Attention Profiling Markup Language (APML) http://www.apml.org/

CALIBRATE: http://calibrate.eun.org

Delicious: http://delicious.com

Diigo (2009) http://www.diigo.com

Flickr: http://www.flickr.com

Learning Resource Exchange: http://lreforschools.eun.org

LeMill: http://lemill.net

MELT: http://info.melt-project.eu

OER Commons: http://www.oercommons.org

User Labor Markup Language (ULML) http://userlabor.org/